Exchange of Spatial and Temporal Resolution in Television Coding

By R. F. W. PEASE and J. O. LIMB

(Manuscript received April 17, 1970)

We describe in this paper a television system in which only a fraction of the picture elements is sampled. A movement detecting circuit examines the frame-to-frame difference signals and divides the picture into "moving" and "stationary" areas. In stationary areas the unsampled elements retain their value from the previous frame; in moving areas the unsampled elements take interpolated values from the current frame. When one-half of the elements is sampled, the resulting pictures are difficult to distinguish from the original, fully sampled, picture. When the fraction is only one quarter, the degradation is visible and disturbing. We mention several possible improvements.

I. INTRODUCTION

Television coders are usually designed to meet simultaneously the worst contingencies with respect to contrast, sharpness and movement. That is, a fast moving subject can be reproduced with the full spatial resolution afforded a stationary subject and with tbe full contrast resolution afforded a low detail subject. By contrast resolution we mean the accuracy with which the coded signal represents tbe amplitude of the input signal for a particular picture element.

We want to reduce the channel capacity required for transmitting television signals hy coding the signal so that full spatial resolution is only available in stationary areas of tbe picture and full temporal resolution is only available in moving areas of the picture.

Exchange of spatial and contrast resolution has been appreciated and demonstrated for some time.[1] One of the first examples of a coder that exchanged spatial and contrast resolution was demonstrated by E. R. Kretzmer[2]; more quantizing levels were assigned to low frequency signal components and fewer levels were assigned to high frequency signal components.

A. J. Seyler[1] has pointed out that this principle can be extended to exchanging temporal and spatial resolution because we can tolerate blurring of moving objects. This is evident because the television camera integrates the video signal on the target for 1/30th of a second. Thus, if the object being imaged on the target moves during this period, the image will be blurred and resolution will be destroyed in the direction in which the object moves. The amount of blurring can be quite large. For example, if the object moves at a speed which would require two seconds to cross the television screen, then in one frame-time a picture element would average light from eight different picture elements on the object (at broadcast television rates). Consequently, if the object is moving in the horizontal direction, the horizontal resolution is reduced approximately eightfold; there is even more reduction if there is camera lag.

It is not clear whether people tolerate such blurring because of the psycho-physical integration of moving objects[3] or because of long exposure to television, but the fact that we do tolerate this loss in resolution means that we should be able to reduce the spatial sampling rate in moving areas of a television picture without degrading picture quality. We refer to sampling at a reduced rate along a scan line (or along a vertical line) as spatial subsampling. The unsampled picture elements are replaced by interpolating between the sampled elements.

If an object moves slowly enough, temporal resolution can be greatly reduced without impairing picture quality. A simple method of achieving this is by frame repeating, in which one new frame of information out of n is transmitted and for the remaining $(n - 1)$ frames this one frame is just repeated.[4] As would be expected if the subject moves fast enough, the perceived motion is very jerky. In alternative, and more pleasing, methods for reducing temporal resolution $1/n^{\text{th}}$ of the points are replenished in each frame in a dot interlaced fashion.[4–6] The type of picture degradation for these schemes differs from frame repeating in that when a subject moves, the edges become blurred and exhibit a checkerboard texture. We refer to sampling a given picture element at the reduced frame rate as temporal subsampling.

By temporally subsampling in stationary areas and spatially subsampling in moving areas we should be able to reduce the channel capacity required to transmit a satisfactory television signal. We have made a digital television system in which the picture is divided into stationary and moving areas which are appropriately subsampled.

This apparatus, the resulting pictures and the saving of required channel capacity are described below.

II. APPARATUS

An outline schematic is shown in Figure 1. The scan format is 30 frames per second with a 2:1 interlace. Each frame has 271 lines, each comprising 248 picture elements. The television signal source is a vidicon camera; the video signal is digitized to 8-bit accuracy in the Pulse Code Modulation (PCM) encoder and stored in the digital ultrasonic wire delay lines of the frame memory. The output of the frame memory is subtracted from the PCM encoder output to yield a frame difference signal for each picture element. These signals are examined in the movement detector whose output determines whether the current element (entering the area) is to be regarded as being in a moving or stationary area. The output changes from a "stationary state" ($s = 0$) to a "moving state" ($s = 1$) when more than n frame difference signals out of a sequence of m exceed a threshold. The output will return to the stationary state when all m samples exhibit insignificant frame difference signals.

Thus the movement detector exhibits hysteresis reducing the frequency of mode changes so that the picture is segmented into relatively few contiguous moving and stationary areas.

In areas judged to be stationary, alternate elements in each line are sampled according to the pattern of Fig. 2a where the elements tabulated A are sampled in one frame and the elements tabulated B are sampled in the next. The unsampled elements retain the value of

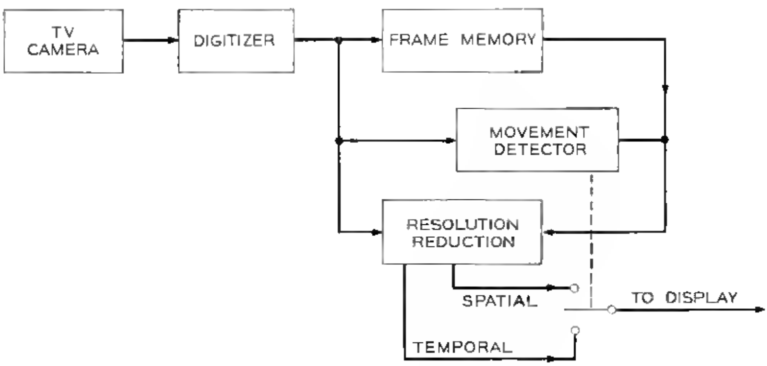

Fig. 1—System for resolution exchange.

the previous frame. In the moving areas the same sampling pattern could be used, but we found that the pattern shown in Fig. 2b gave a better picture; the unsampled elements (marked 0) are given the mean value of the two neighboring elements. When the sampling rate along the line is reduced, we rely on the blurring caused by the motion to bandlimit the video signal. A bandlimiting filter should also be used on the output signal. However, with the sampling pattern of Fig. 2b linear interpolation between the sampled elements gave satisfactory pictures although aliasing patterns were just visible when the subject moved at a certain speed. Further filtering removed these patterns but caused blurring which we judged to be more objectionable.

Strickly speaking, frame-to-frame comb filters should be used in "stationary" areas because of the reduced temporal sampling rate but such filters would involve frame delays and were felt to be impractical. The aforementioned checkerboard texture which can be seen on moving edges in temporally subsampled areas is probably due to the resulting aliasing. The experiments carried out with this apparatus and the resulting pictures are described in the next two sections.

In a second series of experiments only every fourth element was sampled and the sampling patterns for the stationary and moving areas are shown in Figs. 2c and 2d. Only fast movement can now sufficiently bandlimit the video signal and so there are switched filters before and after the subsampling.

Some experiments were also conducted with a 4-bit frame difference quantized signal. The characteristics of the quantizer are:

INPUT LEVELS								
$\pm$	0, 1	2–5	6–11	12–17	18–27	28–37	38–52	54–255
OUTPUT LEVELS								
$\pm$	0	2	8	14	22	32	44	60

III. PROCEDURE

Each scene viewed by the television camera was a head and shoulders view of a person talking; the movement varied from 'gentle' (only lip movement) to 'very active' (arm movement and the person getting up and walking away).

We viewed the resulting pictures on a television monitor. The unblanked raster size was $5\frac{1}{2}$ inches horizontal by 5 inches vertical. The viewing distance was approximately 40 inches and the ambient

			A O O O A O O O A
			A O O O A O O O A
A B A B A B A B A	A O A O A O A O A	A C B D A C B D A	O O A O O O A O O
A B A B A B A B A	A O A O A O A O A	A C B D A C B D A	O O A O O O A O O
B A B A B A B A B	O A O A O A O A O	B D A C B D A C B	A O O O A O O O A
B A B A B A B A B	O A O A O A O A O	B D A C B D A C B	A O O O A O O O A
(a) TEMPORAL 2:1	(b) SPATIAL 2:1	(C) TEMPORAL 4:1	(d) SPATIAL 4:1

Fig. 2—Sampling patterns. ABCD refer to consecutive frames. 0 denotes average of neighboring, sampled elements.

illumination was about the average office illumination (70 foot-candles).

For each scene we first viewed the picture resulting from the 8-bit PCM signal and then the processed pictures in the following order: continuously temporally subsampled; continuously spatially subsampled; temporally subsampled in stationary areas and spatially subsampled in moving areas. As a check we sometimes 'flagged' the area judged moving by the movement detector. For most work the detector was adjusted to switch to spatial subsampling when four out of eight picture elements (pels) examined exhibited frame-to-frame difference signals exceeding a threshold of four (out of 255 levels); for a return to the temporal subsampling mode, none of the eight examined pels exhibited a frame difference exceeding this threshold.

For recording still photographs a model head was swung as a pendulum bob so that the object speed would be known. One television frame (selected at the bottom of the swing) was stored, displayed continuously and photographed with an exposure time of $\frac{1}{4}$ second.

IV. RESULTS

The pictures which were temporally subsampled (by a factor of 2) all over, were not only excellent for stationary scenes but also satisfactory for scenes with an object (e.g., a head) speed up to one pel per frame interval for an object of normal contrast. At higher speeds the checkerboard pattern already referred to and described by others[4] became visible; at object speeds of two pels per frame interval and above, this pattern was annoying for most scenes.

When the pictures were spatially subsampled by a factor of two all over, the loss of resolution was just visible for most scenes; the loss was more obvious when viewing grapbical material.

When the sampling frequency was halved temporally in stationary areas and spatially in moving areas, the resulting pictures were

usually indistinguishable from the fully sampled picture. The scenes where the difference was visible contained a contrasty edge moving slowly (one pel per frame interval or less); in this case the continuity of the edge was disturbed. This effect is caused by erratic movement detection so that parts of the edge are sometimes spatially subsampled and sometimes temporally subsampled. More sophisticated movement detection should reduce this edge breakup by changing modes only when the blurring due to temporal subsampling is comparable witb the blurring due to spatial subsampling.

Some still photographs are shown in Fig. 3. In Fig. 3a the whole scene is stationary and the picture is fully sampled. In Fig. 3b the head is moving at about three pels per frame interval (estimated accuracy $+0, -\frac{1}{2}$ pels per frame interval); the picture is still fully sampled and the blurring introduced by the movement can be seen by comparison with Fig. 3a. Fig. 3c shows an equivalent scene with a 2:1 exchange of resolution. There is very little difference between this picture and the fully sampled one (Fig. 3b). Fig. 3d sbows the bright flags indicating that part of the scene judged to be moving during one frame.

The results using 4:1 subsampling were less encouraging. With overall 4:1 temporal subsampling (according to the pattern of Fig. 2c) the blurring and cbeckerboard patterns were objectionable even in such slowly moving areas as someone's mouth when he was talking. In the pictures resulting from 4:1 spatial subsampling (according to the pattern in Fig. 2d), the blurring was again objectionable, especially in stationary areas.

In the pictures resulting from exchanging resolution according to the movement detector, the breakup of edges was much more apparent and the movement detection was erratic; even some parts of the background or other stationary areas were judged moving. These errors arise because the digital filter in the loop gives rise to frame difference signals wbich may be interpreted as movement by the movement detector. This effect keeps the digital filter switched in even in stationary areas. Again, more sophisticated movement detection should reduce this effect.

The pictures coded as a frame difference quantized signal (instead of 8-bit PCM) were usually similar to those coded as 8-bit PCM. The only differencc was with scenes containing fast movement (speeds of four pels per frame interval or more) of contrasty objects; these objects appeared somewhat noisier than in those pictures coded as eight-bit PCM.

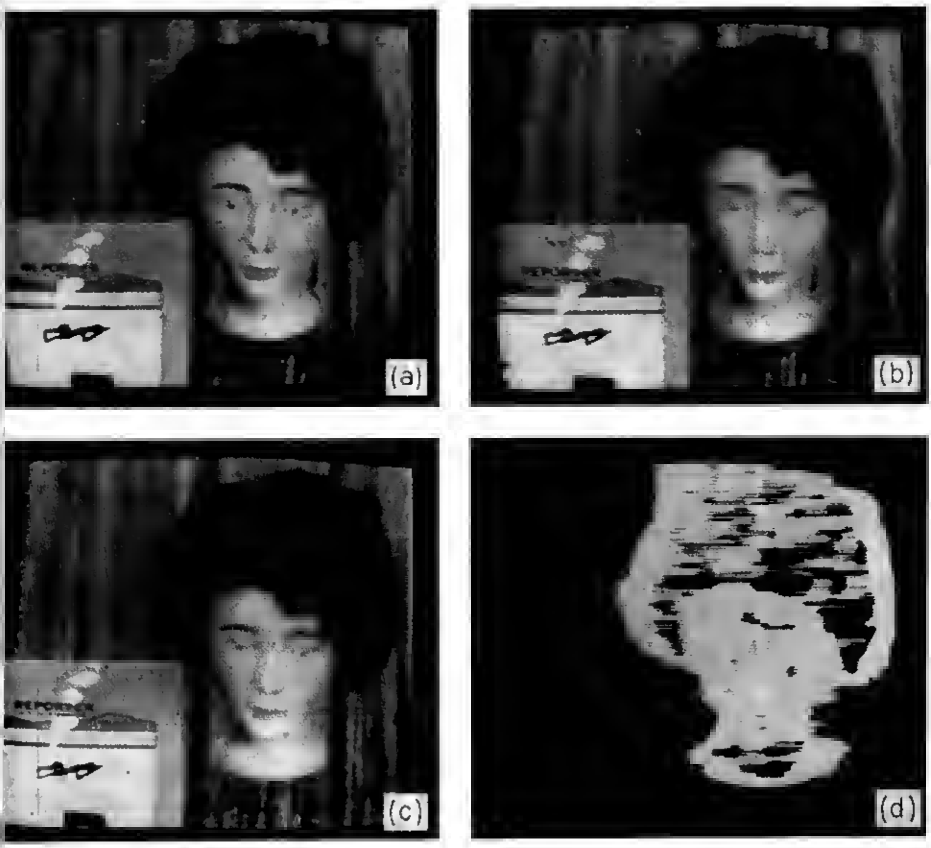

Fig. 3—(a) Fully sampled picture showing stationary head. (b) Fully sampled picture showing moving head. (c) Subsampled (2:1) picture using exchanged resolution. (d) Flags showing area deemed moving. Glossy prints of this figure may be obtained by writing to the authors.

V. DISCUSSION

5.1 *Comparison of 2:1 and 4:1 Subsampling*

Although the simple techniques described above yielded a satisfactory picture when the overall sampling rate was halved, the equivalent techniques did not give a satisfactory picture when the sampling rate was quartered. The unpleasant effect of the latter technique appears mostly at the boundaries between areas treated differently. However, there are many possibilities using a 4:1 ratio reduc-

tion tbat remain to be tried: (*i*) Use an intermediate mode in wbich the second of the three unsampled elements is assigned the value from the previous frame, and the first and tbird are assigned the average value of their neigbbors. Tbis is a simultaneous reduction by a factor of two in both spatial and temporal resolution and sbould result in a less visible boundary between areas encoded in the different modes. (*ii*) Modify the spatial subsampling patterns so that vertical and borizontal resolution is reduced equally. At present tbe horizontal loss is more than the vertical loss. (*iii*) Use an improved movement detector which will indicate the local speed of movement so that modes are not switched until the blurring is similar in each mode.

5.2 *Combination of Subsampling with Other Coding Techniques*

Much digital television employs coders which give either element differential PCM in which the interelement difference along a line is quantized[7] or frame differential PCM in which the difference between consecutive frames is quantized.[8,9]

If the information describing when to switch modes can be transmitted during the blanking interval, then we could use a 4-bit frame difference PCM signal combined with 2:1 resolution exchange for encoding television signals. In our scan format, which resembles that envisaged for *Picturephone*® service, this would require a channel capacity of four megabits per second. We have measured the number of mode changes per line for a number of different scenes and found that it was rarely more than six. If we used a 4-bit word to describe each mode change, then the time required for transmitting the information is rarely more than six microseconds. This is considerably less than the line blanking interval of 13 microseconds for our present scan format. Such a system, although requiring a frame memory, is essentially bufferless as a 6-word shift register suffices to buffer the above mode switching information. Applying spatial subsampling techniques to element differential PCM is not straightforward because wben we halve the horizontal sampling frequency we no longer bave sucb near neighbors for predicting the value of an upcoming element and therefore quantizing noise increases. Two possible alternatives are:

(*i*) Use vertically adjacent elements for predicting the upcoming element. Because of interlace such elements are temporally displaced by one field time from the current line and may be unsatisfactory for predicting values in tbe current line in moving

areas. This problem would not arise with a sequential, non-interlaced, scan format.

(*ii*) Reduce only vertical resolution by sampling alternate lines in each field. Such a process may give rise to an unpleasant rolling pattern in the picture.

We describe experiments in which resolution exchange is applied to an element differential PCM signal in a forthcoming publication.

A third form of coding which is being considered is conditional picture element replenishment[10] in which only the signal corresponding to the moving part of the picture is transmitted. Such a system requires a large buffer store which can become overloaded when the pictures contain large areas of rapid movement. The use of spatial subsampling, either 2:1 or 4:1, in such moving areas is one way of reducing or preventing buffer overflow.[9,11] If we are aiming for a bit rate of one-bit per picture element per frame we can simulate the worst possible case by spatially subsampling by 4:1 the whole picture encoded as 4-bit frame differential PCM. The resulting picture looks much better than the picture tearup which occurs when a buffer overloads[10] and, hence, this mode can be used whenever the buffer approaches overflow.

VI. ACKNOWLEDGMENTS

We would like to acknowledge the many stimulating discussions with A. J. Seyler of the Research Laboratories of the Australian Post Office and with J. C. Candy, B. G. Haskell, and F. W. Mounts.

REFERENCES

1. Seyler, A. J., "The Coding of Visual Signals to Reduce Channel-Capacity Requirements," Proc. IEE, *109,* Part C, No. 16 (September 1962), pp. 676–684.
2. Kretzmer, E. R., "Reduced-Alphabet Representation of Television Signals," 1956 IRE Conv. Rev., Part 4, pp. 140–147.
3. Van den Brink, G., "The Visibility of Details of a Moving Object," Acta Electronica, *2* (1957–58), pp. 44–49.
4. Brainard, R. C., Mounts, F. W., and Prasada, B., "Low Resolution TV: Subjective Effects of Frame Repetition and Picture Replenishment," B.S.T.J., *46,* No. 1 (January 1967), pp. 261–271.
5. Deutsch, S., "Pseudo Random Dot Scan Television Systems," Trans. On Broadcasting IEEE, *BC-11,* No. 1 (July 1965), pp. 11–21.
6. L. C. Jesty, "Television as a Communication Problem," Proc. IEE, *99,* Part 3A, No. 20 (1952), pp. 761–770.
7. Limb, J. O., and Mounts F. W., "Digital Differential Quantizer for Television," B.S.T.J., *48,* No. 7 (September 1969), pp. 2583–2599.
8. Schaphorst, R., "Frame to Frame Coding of N.T.S.C. Color TV," Symposium on Picture Bandwidth Compression, April 1969, M.I.T.
9. Candy, J. C., Franke, M. A., Haskell, B. G., and Mounts, F. W., to be published.

10. Mounts, F. W., "A Video Encoding System Using Conditional Picture-Element Replenishment," B.S.T.J., *48,* No. 7 (September 1969), pp. 2545–2554.
11. Medlin, J. E., "The Prevention of Transmission Buffer Overflow in Telemetry Data Compressors," IEEE Trans. Comm. Tech., *COM-16,* No. 1 (February 1968), pp. 94–107.